UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
10/776,892	02/11/2004	Kirill Stoimenov	9432-000257	3341

27572 7590 11/08/2007
HARNESS, DICKEY & PIERCE, P.L.C.
P.O. BOX 828
BLOOMFIELD HILLS, MI 48303

EXAMINER	
SIEDLER, DOROTHY S	

ART UNIT	PAPER NUMBER
2626	

MAIL DATE	DELIVERY MODE
11/08/2007	PAPER

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

Office Action Summary	Application No.	Applicant(s)	
	10/776,892	STOIMENOV ET AL.	
	Examiner	Art Unit	
	Dorothy Sarah Siedler	2626	

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) OR THIRTY (30) DAYS, WHICHEVER IS LONGER, FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1)☒ Responsive to communication(s) filed on *06 September 2007*.

2a)☒ This action is **FINAL**. 2b)☐ This action is non-final.

3)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

4)☒ Claim(s) *1-29* is/are pending in the application.

4a) Of the above claim(s) _____ is/are withdrawn from consideration.

5)☐ Claim(s) _____ is/are allowed.

6)☒ Claim(s) *1-29* is/are rejected.

7)☐ Claim(s) _____ is/are objected to.

8)☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

9)☐ The specification is objected to by the Examiner.

10)☒ The drawing(s) filed on *11 February 2004* is/are: a)☒ accepted or b)☐ objected to by the Examiner.

Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

11)☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

12)☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

a)☐ All b)☐ Some * c)☐ None of:

1.☐ Certified copies of the priority documents have been received.

2.☐ Certified copies of the priority documents have been received in Application No. _____.

3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

* See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

1) ☒ Notice of References Cited (PTO-892)

2) ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)

3) ☐ Information Disclosure Statement(s) (PTO/SB/08) Paper No(s)/Mail Date _____.

4) ☐ Interview Summary (PTO-413) Paper No(s)/Mail Date. _____ .

5) ☐ Notice of Informal Patent Application

6) ☐ Other: _____.

DETAILED ACTION

This office action is in response to the amendment filled September 6, 2007. Claims 1-29 are pending. Claims 1,10,14-18,22, and 25-28 are amended and claim 29 is added.

Response to Amendment

Applicant has successfully amended claim 10 therefore the 35 U.S.C. §112 second paragraph rejection is withdrawn.

Applicant has successfully amended claims 14-18 and 25-28, therefore the 35 U.S.C. §112 second paragraph rejection is withdrawn.

Response to Arguments

Applicant's arguments filed September 6, 2007 have been fully considered but they are not persuasive. Applicant argues that, " Taylor does not teach, suggest, or motivate communicating variations in one or more types of speech features associated with segments of a processed representation of text-input by varying visual display properties of the segments" (Remarks page 11), however the examiner respectfully disagrees. The examiner notes that the specific limitation recited above, as used in claims 1 and 22, is a recitation of the expected result of the intended use of the claimed invention, and is therefore not given significant patentable weight. Additionally, a recitation of the intended use of the claimed invention must result in a structural difference between the claimed invention and the prior art in order to patentably

invention from the prior art. If the prior art structure is capable of performing the intended use, then it meets the claim.

Secondly, ***Taylor*** does not explicitly disclose using a visual editing interface, however ***Taylor*** does disclose that most SGML documents, such as HTML, are physically typed at keyboards (page 17, Section 5.2, first paragraph). This implies the presence of a word processor, enabling a user to enter the text, including tags, and edit the SGML document. ***Taylor*** also discloses the use of a level tag, which is used to indicate the amount of automatic prosodic analysis initially performed by a machine (pages 12-13, Section 3.4). The level tag enables a user to indicate when the system should automatically produce prosodic tags, and when they should be provided by the user, for example through editing. In edition, ***Henton*** discloses the use of a graphical editor to manipulate the speech output from a text-to-speech system (Abstract). The graphical editor is designed to visually represent and enable the user to control vocal characteristics, such as volume, duration and pitch duration. This visualization creates a faster and more intuitive approach to manipulating speech parameters (page 115-116, section 6).

Applicant's arguments with respect to ***Kobal*** have been considered but are moot in view of the new ground(s) of rejection.

Claim Rejections - 35 USC § 103

The text of those sections of Title 35, U.S. Code not included in this action can be found in a prior Office action.

Claims 1-7,12-19 and 22-28 are rejected under 35 U.S.C. 103(a) as being unpatentable over ***Taylor*** ("SSML: A Speech Synthesis Markup Language" Speech Communication, 1996) in view of ***Henton*** ("Generating and Manipulating Emotional Synthetic Speech on a Personal Computer" 1996).

As per claim 1, ***Taylor*** discloses a system for tuning the text-to-speech conversion process, the system comprising: a text-to-speech engine, said text-to-speech engine receiving at least one text-input and converting said text-input into a processed representation, said processed representation including at least one speech feature associated with at least one segment of said representation (page 3, Section 1.1 Annotated Text in Speech Synthesis, *a markup language is used to annotate, or tag, input text (processed representation), the tags indicating a pronunciation of the input text word or phrase*). ***Taylor*** does not explicitly disclose a visual editing interface, said visual editing interface displaying said processed representation using at least one graphical indicator on an output device, wherein said segment is displayed on said output device using said graphical indicator corresponding to said speech features, thereby communicating variations in one or more types of speech features associated

with segments of said representation by varying visual display properties of the segments. However ***Taylor*** does disclose that most SGML documents, such as HTML, are physically typed at keyboards (page 17, Section 5.2, first paragraph). This implies the presence of a word processor, enabling a user to enter the text, including tags, and edit the SGML document. ***Taylor*** also discloses the use of a level tag, which is used to indicate the amount of automatic prosodic analysis initially performed by a machine (pages 12-13, Section 3.4). The level tag enables a user to indicate when the system should automatically produce prosodic tags, and when they should be provided by the user, for example through editing. In addition, ***Henton*** discloses a graphical user interface, which presents a graphical indicator corresponding to said speech features thus enabling the user to visually represent and adjust speech features (page 115-117, section 6, *a graphical user interface enables the user to visually represent and control vocal characteristics. The color and font size of the word are adjusted, thus adjusting the underlying vocal characteristic (speech feature)*). The user interface enables the user to adjust pronunciation features, including prosody and duration, using a graphical tool. ***Taylor*** and ***Henton*** both disclose systems for the adjustment of prosodic features used during speech synthesis, and are therefore analogous art.

Therefore it would have been obvious to one of ordinary skill in the art at the time of the invention to have a visual editing interface displaying said processed representation using at least one graphical indicator on an output device, wherein said segment is displayed on said output device using said graphical indicator corresponding to said speech feature in ***Taylor***, since a graphical indicator provides a fast and intuitive

method for the user to adjust the vocal characteristics of speech output by the text-to-speech system, as indicated in ***Henton*** (page 115-116, section 6).

As per claim 18, ***Taylor*** does not explicitly disclose a system for providing a text-to-speech interface, the system comprising: a visual interface connected to a text-to-speech engine; and at least one communication channel connecting said visual interface to said text-to-speech engine, said text-to-speech engine communicating with said visual interface over said communication channel by sending and receiving at least one data segment in a format, wherein said visual interface communicates variations in one or more types of speech features associated with segments of said data by varying visual display properties of the segments. However ***Taylor*** does disclose that most SGML documents, such as HTML, are physically typed at keyboards (page 17, Section 5.2, first paragraph). This implies the presence of a word processor, enabling a user to enter the text, including tags, and edit the SGML document. ***Taylor*** also discloses the use of a level tag, which is used to indicate the amount of automatic prosodic analysis initially performed by a machine (pages 12-13, Section 3.4). The level tag enables a user to indicate when the system should automatically produce prosodic tags, and when they should be provided by the user, for example through editing. In addition, ***Henton*** discloses a graphical user interface, which presents a graphical indicator corresponding to said speech features thus enabling the user to visually represent and adjust speech features (page 115-117, section 6, *a graphical user interface enables the user to visually represent and control vocal characteristics. The color and font size of the word*

are adjusted, thus adjusting the underlying vocal characteristic (speech feature)). The user interface enables the user to adjust pronunciation features, including prosody and duration, using a graphical tool. ***Taylor*** and ***Henton*** both disclose systems for the adjustment of prosodic features used during speech synthesis, and are therefore analogous art.

Therefore it would have been obvious to one of ordinary skill in the art at the time of the invention to have a visual interface connected to a text-to-speech engine, and at least one communication channel connecting said visual interface to said text-to-speech engine, said text-to-speech engine communicating with said visual interface over said communication channel by sending and receiving at least one data segment in a format in ***Taylor***, since it would enable a fast and intuitive method for the user to adjust the vocal characteristics of speech output by the text-to-speech system, as indicated in ***Henton*** (page 115-116, section 6).

As per claim 22, ***Taylor*** disclose a method for visual tuning text-to-speech conversion process, the method comprising: converting an input-text to a processed representation using a text-to-speech engine, said processed representation including at least one speech feature of said input-text (page 3, Section 1.1 Annotated Text in Speech Synthesis, *a markup language is used to annotate, or tag, input text (processed representation), the tags indicating a pronunciation of the input text word or phrase*); ***Taylor*** does not explicitly disclose displaying said processed representation on a visual

editing interface connected to said text-to-speech engine, said speech feature of said processed representation being displayed in a corresponding graphical form, and providing an editing function in said visual editing interface to a user for modifying said speech feature in said graphical form, thereby communicating variations in one or more types of speech features associated with segments of said representation by varying visual display properties of the segments. However ***Taylor*** does disclose that most SGML documents, such as HTML, are physically typed at keyboards (page 17, Section 5.2, first paragraph). This implies the presence of a word processor, enabling a user to enter the text, including tags, and edit the SGML document. ***Taylor*** also discloses the use of a level tag, which is used to indicate the amount of automatic prosodic analysis initially performed by a machine (pages 12-13, Section 3.4). The level tag enables a user to indicate when the system should automatically produce prosodic tags, and when they should be provided by the user, for example through editing. In addition, ***Henton*** discloses a graphical user interface, which presents a graphical indicator corresponding to said speech features thus enabling the user to visually represent and adjust speech features (page 115-117, section 6, *a graphical user interface enables the user to visually represent and control vocal characteristics. The color and font size of the word are adjusted, thus adjusting the underlying vocal characteristic (speech feature)*). The user interface enables the user to adjust pronunciation features, including prosody and duration, using a graphical tool. ***Taylor*** and ***Henton*** both disclose systems for the adjustment of prosodic features used during speech synthesis, and are therefore analogous art.

Therefore it would have been obvious to one of ordinary skill in the art at the time of the invention to display said processed representation on a visual editing interface connected to said text-to-speech engine, said speech feature of said processed representation being displayed in a corresponding graphical form, and providing an editing function in said visual editing interface to a user for modifying said speech feature in said graphical form in ***Taylor***, since a graphical indicator provides a fast and intuitive method for the user to adjust the vocal characteristics of speech output by the text-to-speech system, as indicated in ***Henton*** (page 115-116, section 6).

As per claim 2, ***Taylor*** in view of ***Henton*** disclose the system of claim 1, and ***Henton*** further discloses wherein said visual editing interface provides at least one editing function to a user, the editing function enabling the modification of said speech feature associated with said segment through a change in the corresponding said graphical indicator (page 115-119, Figure 2,3 and 4, *a user selects a word and adjusts the duration, volume and prosodics by adjusting the length, height and color of the word*).

Therefore it would have been obvious to one of ordinary skill in the art at the time of the invention to have a visual editing interface that provides at least one editing function to the user, the editing function enabling the modification of said speech feature associated with said segment through a change in the corresponding said graphical indicator in ***Taylor***, since a graphical interface provides a fast and intuitive method for

the user to adjust the vocal characteristics of speech output by the text-to-speech system, as indicated in ***Henton*** (page 115-116, section 6).

As per claim 3, ***Taylor*** in view of ***Henton*** disclose the system of claim 2, and ***Henton*** further discloses a visual editing interface that associates said speech feature corresponding to said segment with said graphical indicator, wherein the user's modification of said graphical indicator results in a corresponding change in said speech feature of said segment (page 115-119, Figure 2,3 and 4, *a user selects a word and adjusts that duration, volume and prosodics by adjusting the length, height and color of the word*).

Therefore it would have been obvious to one of ordinary skill in the art at the time of the invention to a visual editing interface that associates said speech feature corresponding to said segment with said graphical indicator, wherein the user's modification of said graphical indicator results in a corresponding change in said speech feature of said segment in ***Taylor***, since a graphical interface provides a fast and intuitive method for the user to adjust the vocal characteristics of speech output by the text-to-speech system, as indicated in ***Henton*** (page 115-116, section 6).

As per claim 4, ***Taylor*** in view of ***Henton*** disclose the system of claim 1, and ***Taylor*** further discloses wherein said speech feature is at least one of the following: normalized text, part-of-speech, parsing of text, chunking of text, boundary strength, pause

duration, transcription, speech rate, syllable duration, segment duration, pitch, word prominence, emphasis, formant mixing mode, unit selection override, intensity contour, formant trajectories, and allophone rules (page 9-11, Section 3.2 Set of Example Tags, *the Speech Synthesis Markup Language, used to annotate the text, includes tags indicating intonational phrase boundary and emphasis*) .

As per claim 5, ***Taylor*** in view of ***Henton*** disclose the system of claim 1, and ***Henton*** further discloses wherein said graphical indicator comprises at least one of the following: graphical style, font faces, coloring, vertical spacing, horizontal spacing, italicization, boldness, underlining, blinking, crossing-out, text orientation, text rotation, punctuation symbols and graphical symbols (page 115-117, Figure 2,3 and 4, *graphical user interface uses font faces and coloring to enable the user to adjust vocal characteritics*).

Therefore it would have been obvious to one of ordinary skill in the art at the time of the invention to have a graphical indicator use one of the indicators indicated above in ***Taylor***, since they provide a fast and intuitive method for the user to adjust the vocal characteristics of speech output by the text-to-speech system, as indicated in ***Henton*** (page 115-116, section 6).

As per claims 6 and 19, ***Taylor*** in view of ***Henton*** disclose the system of claims 1 and 18, and ***Taylor*** further discloses wherein said processed representation employs a

parameterized aligned sound records format (page 19 and 20, *examples of the SSML tags and text used are provided, which are equivalent to the format style of parameterized aligned sound records format*).

As per claim 7, ***Taylor*** in view of ***Henton*** disclose the system of claim 1, and ***Taylor*** further discloses wherein said segment comprises at least one of the following: word, letter, syllable, pause, word boundary and punctuation-mark (Page 9-11, *tags are used in association with a phrase or word*).

As per claims 12 and 23, ***Taylor*** in view of ***Henton*** disclose the system of claims 1 and 22, and ***Henton*** further discloses wherein said visual editing interface provides the user with speech audio output of said processed representation (Abstract, *the graphical user interface enables the user to simulate and adjust speech to be output with the text-to-speech system*).

Therefore it would have been obvious to one of ordinary skill in the art at the time of the invention to have the visual display provide the user with speech audio output in ***Taylor***, since speech playback coordinated with the corresponding text displayed on the screen enables the user to quickly and accurately adjust the system output in real time.

As per claims 14,15,25 and 26, ***Taylor*** in view of ***Henton*** disclose the system of claims 1 and 22, and ***Taylor*** further discloses wherein the said processed representation is a

modified textual representation of the processed input-text, wherein the said input-text is used to generate said processed representation (page 9, section 3.2 Set of Example Tags and page 19, Example SSML Documents, *input text is used to create SSML text documents*).

As per claims 13,16, 24 and 27 ***Taylor*** in view of ***Henton*** disclose the system of claims 1 and 15, however neither disclose wherein visual editing interface is connected to a data-store for storing and retrieving said representation. However, data storage, for example in the form of a hard drive or removable memory, is often used by computer systems to store processed information. That information can be accessed later for review or further processing. For example, if a user is performing a specific processing task which is suddenly interrupted, the user can store the information gathered thus far, and access it at a later time to continue processing.

Therefore it would have been obvious to one of ordinary skill in the art at the time of the invention to have a visual editing machine connected to a data-store for storing and retrieving said presentation in ***Taylor***, since it would enable the user to store information, and access it a later time for review or further processing.

As per claims 17 and 28, ***Taylor*** in view of ***Henton*** disclose the system of claims 14 and 25, and ***Taylor*** further discloses wherein said modified textual representation is used to generate synthesized speech using a TTS system distinct from said text-to-

speech engine (page 14, sections 4.2 SSML Interpreter and 4.3. Synthesizer Operation, and Figure 2, *an SSML document is created, then passed to a synthesizer which outputs synthesized audio*).

As per claim 29, ***Taylor*** in view of ***Henton*** disclose the system of claim 1, and ***Henton*** further discloses wherein said visual editing interface displays a modified textual representation of said text-input, and variations in visual display for communicating different speech features individually associated with different textual segments of the textual representation include a combination of at least two of: (a) variations in graphical length of the textual segments (page 116-119, section 6, Figure 2,3,4); (b) variations in vertical positions of the textual segments; (c) variations in horizontal spacing of the textual segments; (d) variations in font faces of the textual segments; (page 116-119, section 6, Figure 2,3,4) (e) variations in coloring of the textual segments; (page 116-119, section 6, Figure 2,3,4) (f) variations in styles of the textual segments; (g) variations in orientation of the textual segments; (h) variations in rotation of th textual segments; or (i) punctuation of the textual segments.

Therefore it would have been obvious to one of ordinary skill in the art at the time of the invention to have the visual display communicate different speech features using a combination of textural representations in ***Taylor***, since it would enable a fast and intuitive method for the user to adjust the vocal characteristics of speech output by the text-to-speech system, as indicated in ***Henton*** (page 115-116, section 6).

Claims 8-11,20 and 21 are rejected under 35 U.S.C. 103(a) as being unpatentable over ***Taylor*** in view of ***Henton*** as applied to claims 1 and 18 above, and further in view of ***Kobol*** (7,099,828).

As per claims 8 and 9, ***Taylor*** in view of ***Henton*** disclose the system of claim 1, however neither explicitly disclose wherein said visual editing interface operates as a plug-in for a graphical user interface, wherein said plug-in is an ActiveX control. However, ***Kobol*** discloses a user interface for word pronunciation composition, and indicates that the system can be used as a standalone tool, or can be included in a larger application (column 3 lines 46-49). In addition, Active-x controls were developed in the 1990's by Microsoft to enable enhanced formatting of web pages. Using the standard HTML <object> tags, Active-x enables the users to specify data to control, and how to control it, enabling the web a page to behave more like a program than static pages. Therefore the examiner argues that the use of Active-x controls are old and well known.

Therefore it would have been obvious to one of ordinary skill in the art at the time of the invention to have a visual editing interface that operates as a plug-in control for a GUI, wherein the plug-in is an Active-X control in ***Taylor*** and ***Henton***, since one of ordinary skill in the art has good reason to pursue the options within his or her technical

grasp in order to achieve the predictable result of creating an efficient and reliable user interface.

As per claim 10, ***Taylor*** in view of ***Henton*** disclose the system of claim 1, however ***Taylor*** does not explicitly disclose wherein said visual editing interface allows definition of said input-text by providing a set of text messages containing non-editable text and editable blank slots into which at least part of said input-text can be entered. ***Taylor*** does disclose that most SGML documents, such as HTML, are physically typed at keyboards (page 17, Section 5.2, first paragraph). This implies the presence of a word processor, enabling a user to enter the text, including tags, and edit the SGML document. ***Taylor*** also discloses the use of a level tag, which is used to indicate the amount of automatic prosodic analysis initially performed by a machine (pages 12-13, Section 3.4). The level tag enables a user to indicate when the system should automatically produce prosodic tags, and when they should be provided by the user, for example through editing In addition, ***Henton*** discloses a graphical user interface, which presents a graphical indicator corresponding to said speech features thus enabling the user to visually represent and adjust speech features (page 115-117, section 6, *a graphical user interface enables the user to visually represent and control vocal characteristics. The color and font size of the word are adjusted, thus adjusting the underlying vocal characteristic (speech feature)*). ***Henton*** enables the user to enter speech to be edited, and/or add 'pitch marks' to the text to adjust pitch controls (page 119 section 6.3). ***Henton*** does not disclose non-editable text. However, speech

synthesis systems are commonly used in coordination with many other systems including spoken dialogue, machine translation and voice response systems. In voice response systems, synthesized responses are issued to a user to provide instruction or feedback. A voice response system designer may enable adjustment of voice characteristics of the synthesized responses, but disable a change (use non-editable text) in the text of the commands themselves.

Therefore it would have been obvious to one of ordinary skill in the art at the time of the invention to have said visual editing interface allow definition of said input-text by providing a set of text messages containing non-editable text and editable blank slots into which at least part of said input-text can be entered in ***Taylor*** and ***Henton***, since a graphical indicator provides a fast and intuitive method for a user to adjust the vocal characteristics of speech output by the text-to-speech system, as indicated in ***Henton*** (page 115-116, section 6). In addition, non-editable text would enable the user to adjust vocal characteristics without changing the text, or specific word, of the synthesized output.

As per claim 11, ***Taylor*** in view of ***Henton*** disclose the system of claim 1, however neither disclose wherein said visual editing interface is language independent. ***Kobol*** further discloses wherein said visual editing interface is language independent (column 5 lines 60-67).

Therefore it would have been obvious to one of ordinary skill in the art at the time of the invention to have the visual interface be language independent in ***Taylor*** and ***Henton***, since it would enable the system to be used for applications in more than one language, as indicated in ***Kobol*** (column 3 lines 30-32).

As per claim 20, ***Taylor*** in view of ***Henton*** disclose the system of claim 18, however ***Taylor*** does not disclose wherein said text-to-speech engine sends said data segment in the parameterized aligned sound records format to said visual interface, said visual interface rendering said data segment in a visual form, said visual interface allowing editing of said data segment to produce an edited data segment, said visual interface sending said edited data segment to said text-to-speech engine. However ***Taylor*** does disclose that most SGML documents, such as HTML, are physically typed at keyboards (page 17, Section 5.2, first paragraph). This implies the presence of a word processor, enabling a user to enter the text, including tags, and edit the SGML document. ***Taylor*** also discloses the use of a level tag, which is used to indicate the amount of automatic prosodic analysis initially performed by a machine (pages 12-13, Section 3.4). ***Taylor*** further discloses wherein said processed representation employs a parameterized aligned sound records format (page 19 and 20, *examples of the SSML tags and text used are provided, which are equivalent to the format style of parameterized aligned sound records format* The level tag enables a user to indicate when the system should automatically produce prosodic tags, and when they should be provided by the user, for example through editing. ***Henton*** discloses wherein said visual editing interface

provides at least one editing function to a user, the editing function enabling the modification of said speech feature associated with said segment through a change in the corresponding said graphical indicator (page 115-119, Figure 2,3 and 4, *a user selects a word and adjusts that duration, volume and prosodics by adjusting the length, height and color of the word*). In addition, ***Kobal*** discloses a user interface for word pronunciation composition that communicates with a pronunciation processor to send and receive data, including pronunciation information (column 4 lines 11-33 and Figure 1).

Therefore it would have been obvious to one of ordinary skill in the art at the time of the invention to have a text-to-speech engine that sends said data segment in the parameterized aligned sound records format to said visual interface, said visual interface rendering said data segment in a visual form, said visual interface allowing editing of said data segment to produce an edited data segment, said visual interface sending said edited data segment to said text-to-speech engine in ***Taylor***, since a graphical indicator provides a fast and intuitive method for a user to adjust the vocal characteristics of speech output by the text-to-speech system, as indicated in ***Henton*** (page 115-116, section 6) and ***Kobol*** (column 1 lines 50-60).

As per claim 21, ***Taylor*** in view of ***Henton*** disclose the system of claim 18, however neither explicitly disclose wherein said visual interface sends data to said text-to-speech engine over a first said communication channel and said text-to-speech engine sends data to said visual interface over a second said communication channel. ***Kobol***

discloses a graphical user interface, which communicates with a pronunciation processor to send and receive data (column 4 lines 11-33 and Figure 1).

Therefore it would have been obvious to one of ordinary skill in the art at the time of the invention to have a visual interface send data to said text-to-speech engine over a first said communication channel and have said text-to-speech engine send data to said visual interface over a second said communication channel in ***Taylor*** and ***Henton,*** since separate communication channels insure accurate data transfer without interference from other data types.

Conclusion

Applicant's amendment necessitated the new ground(s) of rejection presented in this Office action. Accordingly, **THIS ACTION IS MADE FINAL**. See MPEP § 706.07(a). Applicant is reminded of the extension of time policy as set forth in 37 CFR 1.136(a).

A shortened statutory period for reply to this final action is set to expire THREE MONTHS from the mailing date of this action. In the event a first reply is filed within TWO MONTHS of the mailing date of this final action and the advisory action is not mailed until after the end of the THREE-MONTH shortened statutory period, then the shortened statutory period will expire on the date the advisory action is mailed, and any extension fee pursuant to 37 CFR 1.136(a) will be calculated from the mailing date of the advisory action. In no event, however, will the statutory period for reply expire later than SIX MONTHS from the date of this final action.

Any inquiry concerning this communication or earlier communications from the examiner should be directed to Dorothy Sarah Siedler whose telephone number is 571-270-1067. The examiner can normally be reached on Mon-Thur 9:30am-5:30pm.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Richemond Dorvil can be reached on 571-272-7602. The fax phone number for the organization where this application or proceeding is assigned is 571-273-8300.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see http://pair-direct.uspto.gov. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free). If you would like assistance from a USPTO Customer Service Representative or access to the automated information system, call 800-786-9199 (IN USA OR CANADA) or 571-272-1000.

DSS

TALIVALDIS IVARS ŠMITS
PRIMARY EXAMINER